

# Introduction to Bioinformatics Resources for NCI CCR Scientists

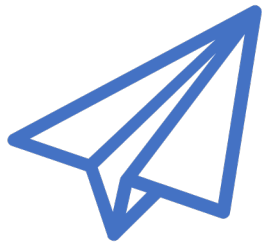
Bioinformatics Training and Education Program (BTEP)

<https://btep.ccr.cancer.gov>

Email: [ncibtep@nih.gov](mailto:ncibtep@nih.gov)

Slides and recording from this presentation  
(<https://btep.ccr.cancer.gov/on-line-classes-2022>)

# We want to hear from you



[Email: ncibtep@nih.gov](mailto:ncibtep@nih.gov)



What kind of training is helpful to you?

# Genome Analysis Unit (GAU) and BTEP Teams

---

Peter Fitzgerald

---

Carl McIntosh

---

Des Tillo

---

Amy Stonelake

---

Joe Wu

---

Alex Emmons

In this  
presentation,  
we will cover



Bioinformatics resources offered by the NCI CCR  
Bioinformatics Training and Education Program (BTEP)



Software purchased by OSTR for CCR researchers



NIH high performance computing Unix cluster  
Biowulf/Helix



NCI Cloud Resources



Resources offered by other NIH training programs



Topic: Resources on the  
BTEP website

# Bioinformatics Training and Education Program



NIH Bioinformatics Calendar at  
<https://btep.ccr.cancer.gov>



Training – Online Classes 2022,  
Upcoming Classes, Class Archive



Resources – Bioinformatics FAQs, Video  
Archive, Class Documentation, New  
Resource Pages



Contact BTEP: [ncibtep@nih.gov](mailto:ncibtep@nih.gov)

# Bioinformatics Training and Education Program

[ncibtep@nih.gov](mailto:ncibtep@nih.gov)
[Home](#) » [Bioinformatics Training and Education Program](#)

## Bioinformatics Training and Education Program

The goal of the Bioinformatics Training and Education Program (BTEP), established by the Office of Science and Technology Resources (OSTR), is to increase the awareness and understanding of bioinformatics techniques and processes among CCR scientists, and to empower CCR scientists to perform a basic, informed set of analyses on their own behalf.

These training classes address the practical aspects of acquiring, representing, and analyzing complex biomedical data sets. Relevant theory and experimental design are also discussed. The goal is to provide both analytic recipes and to ensure better understanding of the theoretical and practical aspects of sound analytical technique.

[REMOTE LEARNING](#)
[VIEW UPCOMING CLASSES](#)

## NIH Bioinformatics Calendar – Upcoming events

Hints on using the calendar, [a grid view of the calendar](#) and [a view with past events](#) is also available.



[TOPIC](#)
[FORMAT](#)
[EVENT ORGANIZER](#)
[PAST & FUTURE EVENTS](#)

### JULY

27  
JUL

**A COMPARATIVE ANALYSIS OF THE MOLECULAR CHARACTERISTICS OF CANINE AND HUMAN GLIOMAS**

28  
JUL

**NCI IMAGING DATA COMMONS, PART OF THE CANCER RESEARCH DATA COMMONS**

### EVENTS AND RESOURCES



**Qlucore Bioinformatics Software for Next Gen Seq Analysis**

Jul 13, 2022 | [Announcement](#)



**Partek Flow: Bulk and Single Cell RNA-Seq Data Analysis**

Jul 13, 2022 | [Resources](#)



**Qiagen IPA Pathway Analysis Online Webinars in July**

Jul 7, 2022 | [Announcement](#)



**Learn Bioinformatics Skills with Dataquest and Coursera! Licenses are available.**

May 31, 2022 | [Resources](#)



**Updated Information on NIDAP bulk and single cell RNA-Seq tutorials**

May 23, 2022 | [Resources](#)

### EXTERNAL LINKS

[Center for Cancer Research](#)

[Office of Science and Technology Resources](#)

[NIH Library Bioinformatics](#)

TOPIC

FORMAT

EVENT ORGANIZER

PAST & FUTURE EVENTS

## JULY

**27**

JUL

**A COMPARATIVE ANALYSIS OF THE MOLECULAR  
CHARACTERISTICS OF CANINE AND HUMAN GLIOMAS**

**28**

JUL

**NCI IMAGING DATA COMMONS, PART OF THE CANCER  
RESEARCH DATA COMMONS**

## AUGUST

**03**

AUG

**CHIP SEQUENCING DATA ANALYSIS**

28  
JUL

## NCI IMAGING DATA COMMONS, PART OF THE CANCER RESEARCH DATA COMMONS



[CALENDAR](#) [GOOGLECAL](#)



### EVENT DETAILS

This seminar series showcases research from different Cancer Moonshot initiatives that support the 10 recommendations of the [Blue Ribbon Panel](#) report. These presentations will inform the community about the progress of Cancer Moonshot-funded projects, provide outreach related to Cancer Moonshot projects, enhance discussions and collaborations related to Cancer Moonshot research, and promote the sharing of data from Cancer Moonshot initiatives.

Speaker: [Andrey Fedorov, Ph.D.](#)[Exit Disclaimer](#), Brigham and Women's Hospital

[more](#) ▼



### REGISTER HERE

REGISTER



### TIME

*(Thursday) 12:00 pm - 1:00 pm*



### LOCATION

*Online*



### ORGANIZER

Cancer Moonshot



# Apply for a Dataquest license

<https://www.surveymonkey.com/r/2NPCBKT>

Become a data expert faster  
with **Dataquest**.



## Learn efficiently

Learn exactly what you need to achieve your goal — and nothing extra.



## Challenge yourself with exercises

Work with real data from day one with hands-on exercises.



## Build your project portfolio

Gain confidence and show off your data skills with projects.

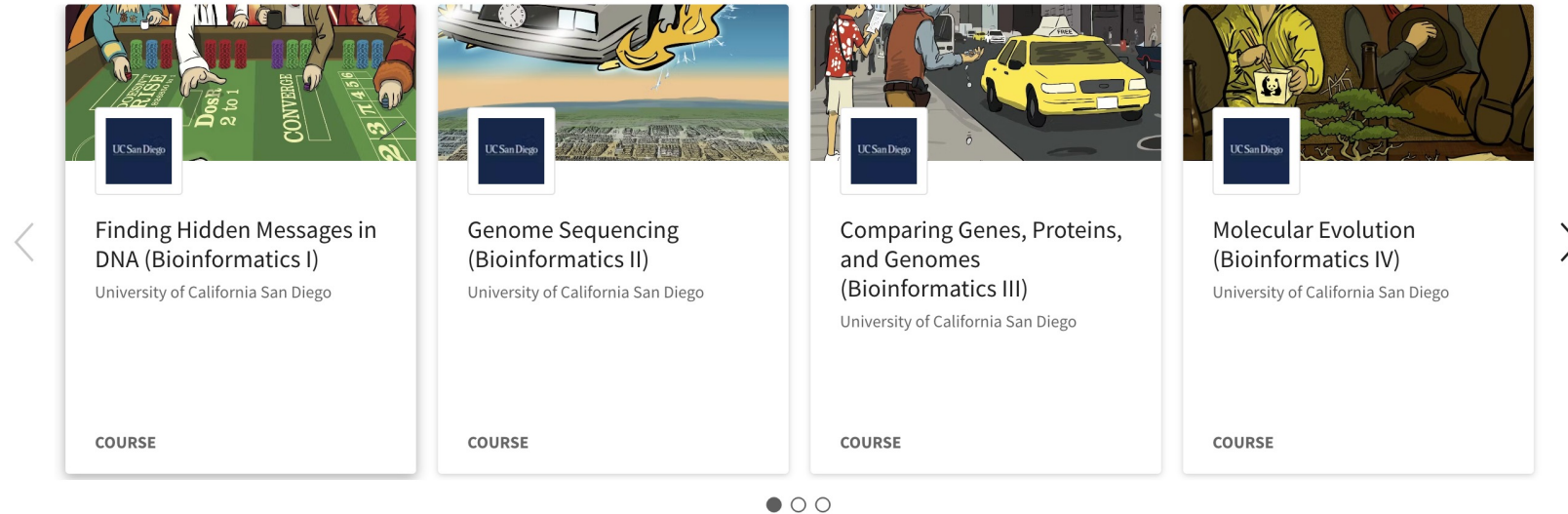


```
1 # Try a Dataquest exercise!
2 # We loaded in the Fortune 500 list
3 # Hit Run Code to show average revenue
4
5 avg_revenue = mean(revenues)
6 print(f"Average revenue: ${avg_revenue}M")
7
```

Output

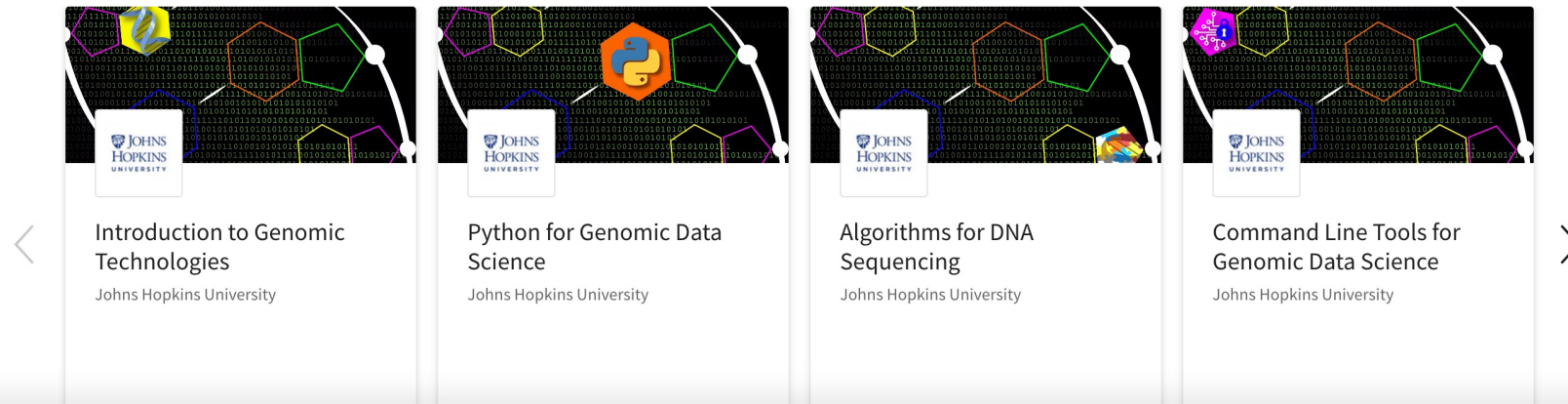
Run Code

Apply for NIH Coursera license [here](#)  
Contact [ncibtep@nih.gov](mailto:ncibtep@nih.gov)



#### Genomic Data Science

This collection has been recommended by your organization.



# Courses so far in 2022



Distinguished Speakers Seminar Series  
Rahul Satija, Melissa Haendel, Sarah  
Teichmann, Nicholas Navin



Topics in Bioinformatics – Microbiome  
Analysis with QIIME2, Variant Analysis



R classes – R basics, Visualization, and Data  
Wrangling



Software how-to – Bulk and Single Cell  
RNA-Seq, Pathway Analysis, Variant  
Analysis, Analyzing Publicly Available Data

# BTEP Distinguished Speakers Seminar Series



**Feb. 17, 1 PM.**

Title: Integrated Analysis of Single Cell Data Across Technologies and Modalities

Rahul Satija, D.Phil., Core Faculty Member, New York Genome Center, Associate Professor of Biology, Center for Genomics and Systems Biology, New York University (NYU), Associate Faculty, Institute for Systems Genetics, NYU Langone Medical Center



**April 21, 1 PM (Recording)**

Title: Realizing Data Interoperability Across Basic Research, Clinical Care, and Patients

Melissa Haendel, Ph.D., Professor and Chief Research Informatics Officer, Marsico Chair in Data Science, University of Colorado, Anschutz Medical Campus



**June 16, 1 PM.**

Title: Mapping the Human Body One Cell at a Time

Sarah Teichmann, Ph.D., Fellow of the Academy of Medical Sciences (UK FMedSci), Fellow of the Royal Society (FRS), Wellcome Sanger Institute

WebEx recording: [BTEP DSSS Sarah Teichmann-20220616 1704-1](#)



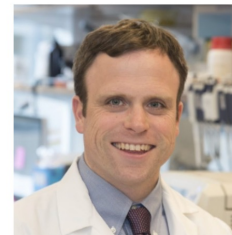
**July 14, 1 PM.**

Title: Decoding Breast Cancer Progression with Single Cell Genomics

Nicholas Navin, Ph.D., Professor, MD Anderson Cancer Center, joint appointment in Dept of Bioinformatics Faculty Member, The University of Texas, MD Anderson Cancer Center, UT Health Graduate School of Biomedical Sciences Director of the CPRIT Single Cell Genomic Center Co-Director, Advanced Technology Genomics Core at MD Anderson

[Register Here](#)

[Meeting Link](#)



**Sept 22, 1 PM**

Christopher E. Mason, Ph.D., Professor, Department of Physiology and Biophysics Weill Cornell Medicine Director, WorldQuant Initiative for Quantitative Prediction

[Register Here](#)

[Meeting Link](#)



**October 27, 1 PM**

Christina Curtis, Ph.D., M.Sc, Associate Professor of Medicine and Genetics, Stanford University School of Medicine, and Co-Director, Molecular Tumor Board, Stanford Cancer Institute

[Register Here](#)

[Meeting Link](#)

# BTEP Online Classes 2022

## (<https://btep.ccr.cancer.gov/on-line-classes-2022/>)

---

### Partek Flow: Bulk and Single Cell Gene Expression Visualization (May 18)

Recording link: <https://cbiit.webex.com/cbiit/ldr.php?RCID=b5b1948aab3093124275d06fa55b08ae>

### Qlucore: Import and analyze public data from SRA, GEO and TCGA (May 11)

Recording link: <https://cbiit.webex.com/cbiit/ldr.php?RCID=c8e8e8ff837caeb2d4d3839bc180b9bb>

### Training: Access GEO, SRA, ArrayExpress, TCGA, GTEx and more with Qiagen IPA Land Explorer (April 20)

Recording link: <https://cbiit.webex.com/cbiit/ldr.php?RCID=d081980970fce4cb50dd20d3f4f3e172>

### Single Cell RNA-Seq Analysis with Partek Flow (April 13)

Recording Link:  
<https://cbiit.webex.com/cbiit/ldr.php%3FRCID%3D02b734d9f2b8466bf996ccc5d52d0c02>

### Qlucore: Pathway Analysis with Gene Set Enrichment Analysis (GSEA) (April 6)

Recording link: <https://cbiit.webex.com/cbiit/ldr.php?RCID=8cfdd939b601f0402b312ba2f8e85262>

Recommended Resources:

1. GSEA hands-on webinar\_Qlucore.pdf
2. GSEA in Qlucore.pdf

### Variant Analysis – Experimental Design, Best Practices, and Workflows (June 30)

Recording link:  
<https://cbiit.webex.com/recordingservice/sites/cbiit/recording/8cd6adf4dacc103a83be00505681c451>

Presentation slides:

Variant analysis – presentation 1  
Variant analysis – presentation 2

Chat history: [chat history](#)

### Qlucore: Bulk RNA-Seq Data Analysis (June 29)

Recording link:  
<https://cbiit.webex.com/webappng/sites/cbiit/recording/c788382fd9e1103ab7f6005056818fce>

### Data Wrangling with R (June 7 – July 7)









Course Materials: <https://btep.ccr.cancer.gov/docs/data-wrangle-with-r/>






**Lesson 1, June, 7th:** Introduction to R, RStudio, and the Tidyverse

This will be a no coding introduction to R, RStudio, and the Tidyverse. In this lesson, we will review some of the advantages of using R for data analysis and will get you acquainted with the RStudio environment. The help session will be devoted to getting everyone connected to the course on DNAnexus.

Recording link: <https://cbiit.webex.com/cbiit/ldr.php?RCID=5544470024885f1aa54c26fc5c9ff6cf>

# FAQs: Single Cell RNA-Seq, ChIP-Seq Analyses on the BTEP website

	<b>Why don't the scRNA transcript counts match up with the results I found with flow cytometry?</b> kellymc@nih.gov answered 2 years ago • Single-Cell RNA-Seq	1766 views	1 answers	0 votes
	<b>For single nuclei RNA-Seq, do I have to do anything different in handling the data?</b> dadkhahe@nih.gov answered 2 years ago • Single-Cell RNA-Seq	1470 views	1 answers	0 votes
	<b>Difference between 3' and 5' gene expression profiling? Which should I use?</b> dadkhahe@nih.gov answered 2 years ago • Single-Cell RNA-Seq	7710 views	2 answers	0 votes
	<b>What is Cell Ranger?</b> dadkhahe@nih.gov answered 2 years ago • Single-Cell RNA-Seq	1459 views	1 answers	0 votes
	<b>Do I need to have extensive bioinformatic knowledge to explore my Cell Ranger results?</b> dadkhahe@nih.gov answered 2 years ago • Single-Cell RNA-Seq	1444 views	1 answers	0 votes
	<b>Droplet-based (e.g. 10x Genomics) or plated-based (e.g. Smart-Seq)? When should one be considered over the other</b> dadkhahe@nih.gov answered 2 years ago • Single-Cell RNA-Seq	4548 views	1 answers	0 votes
	<b>What sequencing depth do I need for scRNA-Seq samples?</b> dadkhahe@nih.gov answered 2 years ago • Single-Cell RNA-Seq	2727 views	2 answers	0 votes
	<b>How do we decide what clustering resolution to set?</b> kellymc@nih.gov asked 2 years ago • Single-Cell RNA-Seq	1370 views	0 answers	0 votes
	<b>Are there standard pre-processing criteria that should be used?</b> kellymc@nih.gov asked 2 years ago • Single-Cell RNA-Seq	1228 views	0 answers	0 votes

	<b>Should we always use histone markers just to see if our sequencing is working for our samples?</b> stonelakeak@nih.gov answered 2 years ago • ChIP-Seq Data Analysis	1376 views	1 answers	0 votes
	<b>What do you think about ChIP-seq with proteins that are highly mobile in nature? Why is there less data published about using them?</b> stonelakeak@nih.gov answered 2 years ago • ChIP-Seq Data Analysis	1187 views	1 answers	0 votes
	<b>I had great results with ChIP-qPCR, but when I did ChIP-seq only a few reactions came up with decent peaks. What could be the reason?</b> stonelakeak@nih.gov answered 2 years ago • ChIP-Seq Data Analysis	1247 views	1 answers	0 votes
	<b>Comparing to outside sources of data: What is a good website to check for transcription factor binding motifs and how they regulate expression?</b> stonelakeak@nih.gov answered 2 years ago • ChIP-Seq Data Analysis	1232 views	1 answers	0 votes
	<b>Comparing to outside sources of data: Do you have recommendations for frameworks or programs for integrating ChIP-seq and RNA-seq or microarray data?</b> stonelakeak@nih.gov answered 2 years ago • ChIP-Seq Data Analysis	1146 views	1 answers	0 votes
	<b>Library/ChIP preparation: What are nanobodies? Can they be used for ChIP-Seq?</b> stonelakeak@nih.gov answered 2 years ago • ChIP-Seq Data Analysis	1300 views	1 answers	0 votes

# BTEP Resources Pages

Go to the following website to check out the new BTEP bioinformatics resources pages :

<https://btep.ccr.cancer.gov/docs/resources-for-bioinformatics/>

The screenshot shows a web browser window with the address bar displaying `btep.ccr.cancer.gov/docs/resources-for-bioinformatics/`. The page header features the NIH National Cancer Institute logo and the text "Bioinformatics Training and Education Program --- email BTEP at ncibtep@nih.gov". Below the header is a red navigation bar with the text "Bioinformatics Resource for CCR Scientists 2022" and a search bar. The main content area is divided into two columns. The left column contains a sidebar with the title "Bioinformatics Resource for CCR Scientists 2022" and a list of links: "Home", "The Bioinformatics Training and Education Program (BTEP)", "Core Facilities: Data pre-processing and data returning policies", "Biowulf High Performance Computing system", "Transferring Large Files with Globus", and "Bioinformatic interest groups, listservs, and Slack channels". The right column has the title "BTEP - Bioinformatics Resources for CCR Scientists" and a download icon. Below the title, it states: "These pages list and describe the main resources available to CCR scientist for carrying out bioinformatic analysis on their data." and "These resources include:". A bulleted list follows: "• Places to obtain training and assistance - BTEP Resources" and "• Information about data delivered by the NCI sequencing facilities".

Bioinformatics Resource for CCR Scientists 2022

Home

The Bioinformatics Training and Education Program (BTEP)

Core Facilities: Data pre-processing and data returning policies

Biowulf High Performance Computing system

Transferring Large Files with Globus

Bioinformatic interest groups, listservs, and Slack channels

BTEP - Bioinformatics Resources for CCR Scientists

These pages list and describe the main resources available to CCR scientist for carrying out bioinformatic analysis on their data.

These resources include:

- Places to obtain training and assistance - BTEP Resources
- Information about data delivered by the NCI sequencing facilities



**Bioinformatics Resource for CCR Scientists 2022**

[Home](#)

[The Bioinformatics Training and Education Program \(BTEP\)](#)

[Core Facilities: Data pre-processing and data returning policies](#)

[Biowulf High Performance Computing system](#)

[Transferring Large Files with Globus](#)

[Bioinformatic interest groups, listservs, and Slack channels](#)

[More Training Opportunities](#)

[Self Learning Platforms](#) >

[Select Software By Topic](#)

[Non-commercial Software](#) >

[Commercial Software](#) >

## Core Facilities: Data pre-processing and data returning policies



**Table of contents**

[Core Facilities](#)

[Understanding QA/QC reports](#)

[fastqc](#)

[multiqc](#)

### Core Facilities

There are a number of core facilities available to NCI researchers. See more information from the [Office of Science and Technology Resources](#).

We most commonly see data from the following cores:

1. [CCR Sequencing Facility \(CCR-SF\)](#) - located at the ATRF in Frederick, MD. This core is dedicated to high throughput sequencing.

- For large scale projects and production ready projects (compare with NCI CCR Genomics Core)



[Summary of Technologies](#)



2. [NCI CCR Single Cell Analysis Facility \(SCAF\)](#) - located on the NIH Bethesda main campus and provides advanced single-cell genomics technologies.

- Primarily for CCR researchers on the Bethesda campus.



## Bioinformatics Resource for CCR Scientists 2022

Home

The Bioinformatics Training and Education Program (BTEP)

Core Facilities: Data pre-processing and data returning policies

**Biowulf High Performance Computing system**

Transferring Large Files with Globus

Bioinformatic interest groups, listservs, and Slack channels

More Training Opportunities

Self Learning Platforms >

Select Software By Topic

Non-commercial Software >

Commercial Software >

## Getting Started with Biowulf

**Biowulf** is the NIH high performance computing cluster. It is a linux computing cluster with greater than 105,000 processors. The NIH HPC systems also house "hundreds of scientific programs, packages and databases" (<https://hpc.nih.gov/apps/>).

Bioinformatic processes often require a lot of memory and computational time, which is limited on individual (local) computers. For bioinformatics tasks that require a lot of memory or can be run in parallel to reduce the time to completion, consider performing such tasks on Biowulf. To obtain a Biowulf account, see the **Biowulf help pages**. A Biowulf account is accessible to all NIH employees and contractors listed in the NIH Enterprise Directory for a nominal fee of \$35 a month.

## Working on the NIH High Performance Unix Cluster Biowulf

### Logging into Biowulf from MacOS

Find the program "Terminal" on your machine, and enter the following statement at the prompt:

```
ssh username@biowulf.nih.gov
```



## Table of contents

Working on the NIH High Performance Unix Cluster Biowulf

Logging into Biowulf from MacOS

Logging into Biowulf from Windows 10 OS

Working on Biowulf - two things you should always do.

Being a good citizen on Biowulf

Running Interactive Jobs

Batch Jobs

Swarm-ing on Biowulf



**Bioinformatics Resource for  
CCR Scientists 2022**

Home

The Bioinformatics Training and  
Education Program (BTEP)

Core Facilities: Data pre-  
processing and data returning  
policies

Biowulf High Performance  
Computing system

Transferring Large Files with  
Globus

Bioinformatic interest groups,  
listservs, and Slack channels

More Training Opportunities

Self Learning Platforms >

Select Software By Topic

Non-commercial Software >

Commercial Software ✓

...

## NIH Integrated Data Analysis Platform (NIDAP)

**Table of contents**

Description

### Description

**NIDAP (NIH Integrated Data Analysis Platform)** is an innovative, cloud-based, collaborative data aggregation and analysis platform that hosts user-friendly bioinformatics workflows and component analysis and visualization tools developed by the NCI developer community based on open source tools and makes them immediately available to biologist end-users across the Institute. --- (<https://ccbr.ccr.cancer.gov/education-training/nidap-workflows/>)

It is a free resource to NCI researchers that can be used for bulk and single cell RNA-Seq analyses. The platform is a graphic user interface (GUI) that does not require users to read or write code. Free training is offered monthly so that researchers can learn to use the workflows on the platform and understand the results.

For more information, including how to access the platform, and current training dates, please see: [https://btep.ccr.cancer.gov/nidap\\_announce/](https://btep.ccr.cancer.gov/nidap_announce/).

# BTEP – Select Software by Topic

<b>Bioinformatics Resource for CCR Scientists 2022</b>	<b>Commercial Bioinformatics Packages</b>	<b>Table of contents</b>
Home	Molecular Biology	Open Source Bioinformatics Tools
The Bioinformatics Training and Education Program (BTEP)	<a href="#">Sequence comparison</a>	<b>Commercial Bioinformatics Packages</b>
Core Facilities: Data pre-processing and data returning policies	<a href="#">Phylogenetics</a>	Molecular Biology
Biowulf High Performance Computing system	<a href="#">Molecular cloning</a>	Variant Analysis
Transferring Large Files with Globus	<a href="#">Restriction digest</a>	Gene Expression
Bioinformatic interest groups, listservs, and Slack channels	<a href="#">Ligation simulation</a>	Epigenetics
More Training Opportunities	<a href="#">PCR primer design</a>	Metagenomics
Self Learning Platforms >	<a href="#">CRISPR editing</a>	Biological Insights
<b>Select Software By Topic</b>		
Non-commercial Software >		
Commercial Software >		
	Variant Analysis	
	<a href="#">Single nucleotide variants</a>	
	<a href="#">Insertions, deletions</a>	
	<a href="#">Structural variants</a>	
	<a href="#">Low frequency variants</a>	
	<a href="#">Copy number analysis</a>	
	<a href="#">Loss of heterozygosity</a>	

## Bioinformatics Resource for CCR Scientists 2022

[Home](#)

[The Bioinformatics Training and Education Program \(BTEP\)](#)

[Core Facilities: Data pre-processing and data returning policies](#)

[Biowulf High Performance Computing system](#)

[Transferring Large Files with Globus](#)

[Bioinformatic interest groups, listservs, and Slack channels](#)

[More Training Opportunities](#)

[Self Learning Platforms](#) >

[Select Software By Topic](#)

[Non-commercial Software](#) >

[Commercial Software](#) >

## Gene Expression



[Gene expression by microarray](#)



[microRNA by microarray](#)



[RNA sequencing](#)



[microRNA sequencing](#)



[Single cell RNA sequencing](#)



[Spatial transcriptomics](#)



## Epigenetics



[ATAC sequencing](#)



[Single cell ATAC sequencing](#)



[ChIP sequencing](#)



[Methylation array](#)



[Methylation tiling array](#)



[Bisulfite sequencing](#)



## Table of contents

[Open Source Bioinformatics Tools](#)

[Commercial Bioinformatics Packages](#)

[Molecular Biology](#)

[Variant Analysis](#)

[Gene Expression](#)

[Epigenetics](#)

[Metagenomics](#)

[Biological Insights](#)

# Upcoming courses of interest

## Bioinformatics for Beginners

- Unix/ Biowulf
- Bulk RNA-Seq Data Analysis

## Distinguished Speakers Seminar Series

- Christopher Mason
- Christina Curtis

## Topics in bioinformatics

- Bulk RNA-Seq
- Single Cell RNA-Seq
- ChIP-Seq
- Microbiome analysis

Next topic:  
Data Analysis  
Options

Licensed Software

NIH HPC BIOWULF

# How should you analyze your data?

---



Using proprietary, point-and-click software purchased for NCI CCR scientists by Office of Science and Technology Resources (OSTR)

May not always be in an environment where these are available

Partek Flow, and Partek Genomics Suite, Qiagen Ingenuity Pathway Analysis, Qlucore Omics Explorer



Learn open source tools, step-by-step

Lots to learn (Unix, R, Python, HPC Biowulf/Helix)

Knowledge transfers outside of NIH

GATK, GSEA, BWA, RSEM, STAR, Salmon

# Office of Science and Technology Resources (OSTR)



PURCHASES SOFTWARE FOR DATA  
ANALYSIS



MAKES LICENSES TO SOFTWARE  
AVAILABLE TO ALL NCI CCR  
RESEARCHERS (SOME ARE ALSO  
AVAILABLE TO ALL NCI  
RESEARCHERS, NOT JUST CCR)



NIH LIBRARY OFFERS SOME OF THE  
SAME LICENSES FOR ALL OF NIH



TYPES OF DATA ANALYSIS: NEXT  
GEN SEQUENCING, STATISTICS,  
PATHWAY

# NCI CCR OSTR licensed software

Partek Flow and Partek Genomics Suite

Qiagen Ingenuity Pathway Analysis and OmicSoft Land Explorer

Qlucore Omics Explorer

Qiagen CLC Genomics Workbench

SnapGene

LaserGene

Geneious (Prime)

Graph Pad Prism

# Molecular Biology Software

Geneious (Prime)

SnapGene

LaserGene

Partek Genomics Suite

CLC Genomics Workbench

Focus on:

---

Partek Flow for Single  
Cell RNA-Seq Analysis

---

Qiagen IPA for  
Pathway Analysis and  
'omics data access

# Partek Flow Applications

Single Cell RNA-Seq (and spatial transcriptomics)

Gene Expression (RNA-Seq, qPCR, microarray data)

Variant Detection

CNV Analysis

Metagenomics (microbial genomes)

Methylation

ChIP-Seq

Non-coding RNA Analysis

Multimic Analysis

# How to Learn Partek Flow

Extensive library of  
webinars and tutorials  
on [Partek.com](https://partek.com)

BTEP-sponsored  
training sessions,  
recordings and  
resource pages

# Partek Flow: Single Cell RNA-Seq Data Analysis

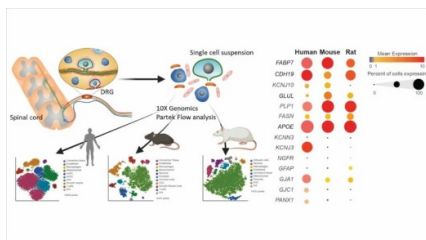
Start with fastq or count matrix (10x Genomics)

- Filtering (exclude barcodes, unique molecular identifiers)
- Drop out low quality cells and uninformative genes
- Interactive QA/QC plots (reads, genes and %mitochondrial reads/cell)

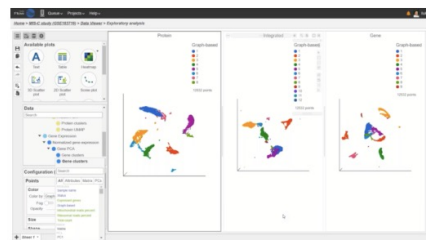
Can do:

- Trajectory analysis
- Differential gene expression analysis
- Tissue transcriptomics

## Our Single Cell Webinar Series



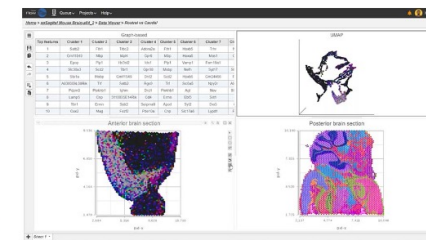
Satellite Glial Cells Role in Neural Repair



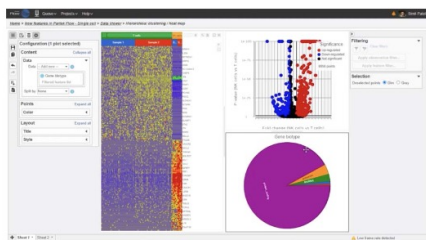
Single Cell Multiomics with COVID-19 Data



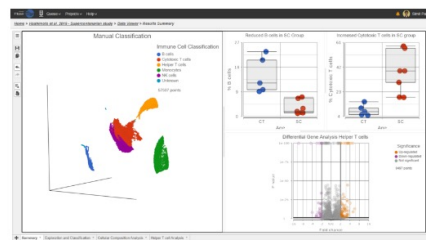
Multomics Analysis



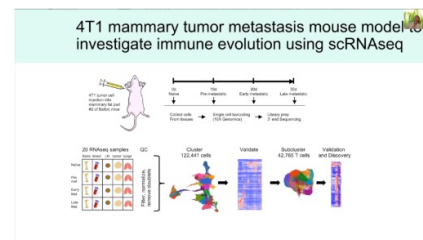
Spatial Transcriptomics Analysis



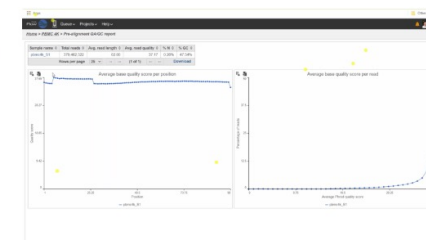
Partek Flow - New Single Cell and Bulk RNA-Seq Features



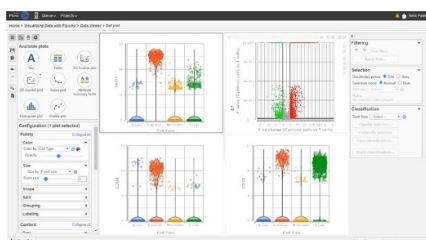
Single Cell RNA-Seq Analysis



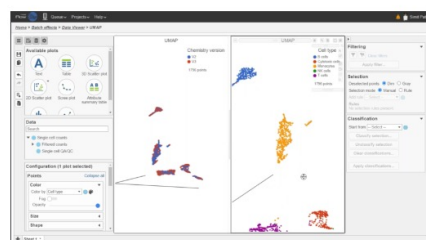
Evolution of Mouse Immune Response During Mammary Tumor Progression and Metastasis



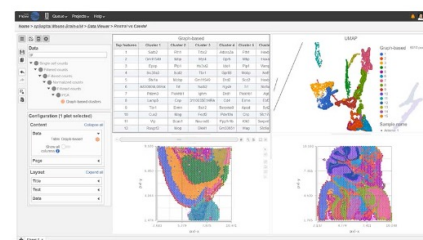
Understanding Data Formats and Preprocessing



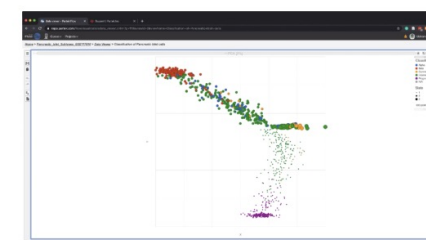
Visualizing Single Cell Data with Figures



Differential Gene Expression Analysis



10x Visium Spatial Gene Expression Data



Single Cell Analysis - inDrop System and Partek Flow

Qiagen  
Ingenuity  
Pathway  
Analysis (IPA)  
and OmicSoft  
Land Explorer

Licenses recently purchased for CCR scientists to access Qiagen OmicSoft Land Explorer which contains data from GEO, SRA, ArrayExpress, TCGA, GTEx and more

In addition to Qiagen IPA licenses for Pathway Analysis

# Qiagen Ingenuity Pathway Analysis

## Understand complex 'omics data to accelerate your research



Discover why QIAGEN Ingenuity Pathway Analysis (IPA) is the leading pathway analysis application among the life science research community and is cited in tens of thousands of articles for the analysis, integration and interpretation of data derived from 'omics experiments. Such experiments include RNA-seq, small RNA-seq, metabolomics, proteomics, microarrays including miRNA and SNP, and small-scale experiments. With QIAGEN IPA you can predict downstream effects and identify new targets or candidate biomarkers. QIAGEN Ingenuity Pathway Analysis helps you perform insightful data analysis and interpretation to understand your experimental results within the context of various biological systems.



# Qiagen OmicSoft Land Explorer

## Contextualize your findings with a massive collection of 'omics data

QIAGEN IPA now offers contextually relevant links to QIAGEN OmicSoft Land Explorer, a database of curated disease 'omics data with over 500,000 samples. With this new premium tier of IPA, you can explore 'omics data for individual genes as well as expression correlation across multiple genes with robust visualizations. Jump from a gene of interest in QIAGEN IPA to discover its tissue or cell expression. Explore the diseases and treatment contexts in which it is up-or-down-regulated. Visualize how mutations correlate with changes in expression, the effect of mutations on clinical outcomes and much more. Access to OmicSoft Land Explorer through IPA requires additional licensing.

Further details on the integration of QIAGEN IPA with QIAGEN OmicSoft Land Explorer can be [found here](#).



Topic: NIH HPC Biowulf

# Biowulf (high-performance cluster)



[hpc.nih.gov](https://hpc.nih.gov)



Thousands of analysis tools (modules) maintained by staff



Scientific reference databases



Next-gen sequencing, computational chemistry, math, statistics, image analysis



User guides and training classes



Monthly Zoom-In Consults see BTEP NIH Calendar

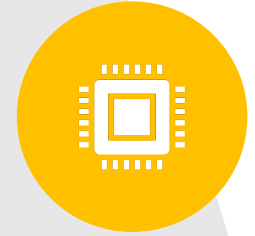
# Why you should get to know Biowulf



BIOWULF IS THE HIGH  
PERFORMANCE CLUSTER  
(HPC) AT NIH.



IT CAN HOLD A LOT MORE  
DATA THAN YOUR PERSONAL  
COMPUTER.



IT HAS MUCH MORE  
COMPUTE RESOURCES THAN  
YOUR PERSONAL  
COMPUTER.



IT CAN HELP YOU ANALYZE  
"BIG DATA".



IT IS AVAILABLE TO ALL NCI  
CCR RESEARCHERS (ALL NIH  
RESEARCHERS).

# Connecting to Biowulf

You can log onto Biowulf from  
MacOS or Windows PC

You'll need to learn some Unix/  
command-line/ shell

Several different ways to use  
compute resources (batch, swarm)

# How to learn more about Biowulf



BTEP Biowulf Beginner Classes



HPC Biowulf Monthly Zoom-In  
Consults (check BTEP calendar)



[hpc.nih.gov](https://hpc.nih.gov)

# Data Transfer and Sharing on HPC Biowulf

Globus can transfer very large data files between your machine and Biowulf

You can mount a Biowulf drive on your local machine

Transfer to and from Cloud resources

Share data with collaborators (inside and outside of NIH)

(<https://hpc.nih.gov/docs/transfer.html>)

# NCI and NIAID Workflows on Biowulf

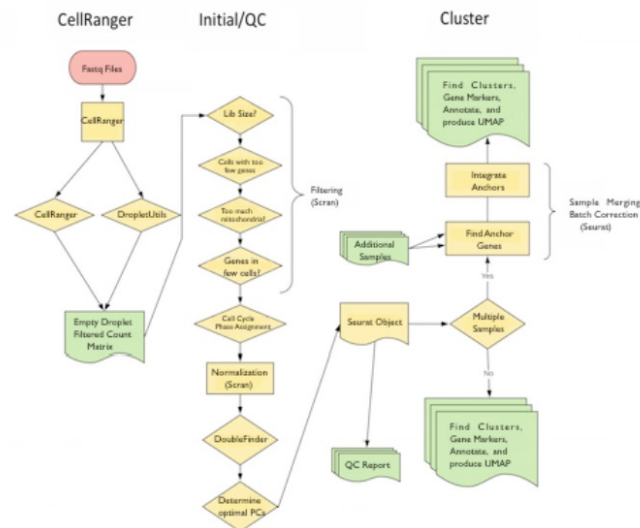
CCBR Pipeliner – bulk and single cell RNA-Seq and ChIP-Seq analysis workflows on NIH HPC Biowulf from the CCR Collaborative Bioinformatics Resource (<https://github.com/CCBR>)

NIAID Collective Bioinformatics Resource (NCBR) OpenOmics/genome-seek: Clinical Whole Genome Sequencing Pipeline (<https://github.com/OpenOmics/genome-seek>)

# Single Cell RNA-Seq Pipeline

## CCBR Pipeliner and Software

- Whole Exome and Genome pipelines
- Single Cell RNA-Seq pipeline
- RNA-Seq Pipeline
- ChIP-Seq Pipeline



### Starting points

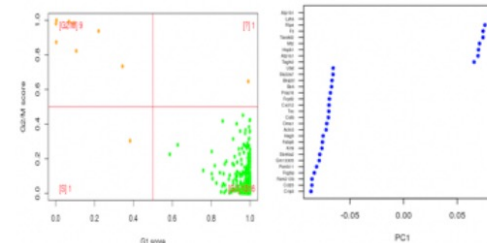
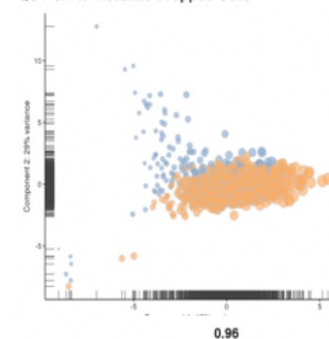
- 10X genomics fastq
- 10X genomics count matrix

### Data Filtering and QC

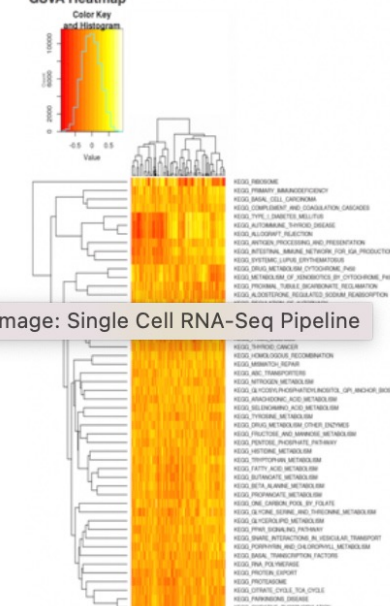
### Downstream analysis

- k-means clustering
- PCA
- tSNE plot
- marker gene lists

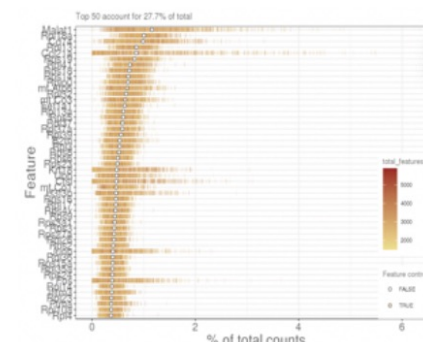
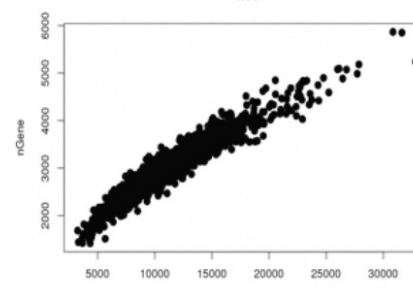
QC PCA to Visualize Dropped Cells



GSEA Heatmap



View image: Single Cell RNA-Seq Pipeline





Next topic: Cloud  
Resources

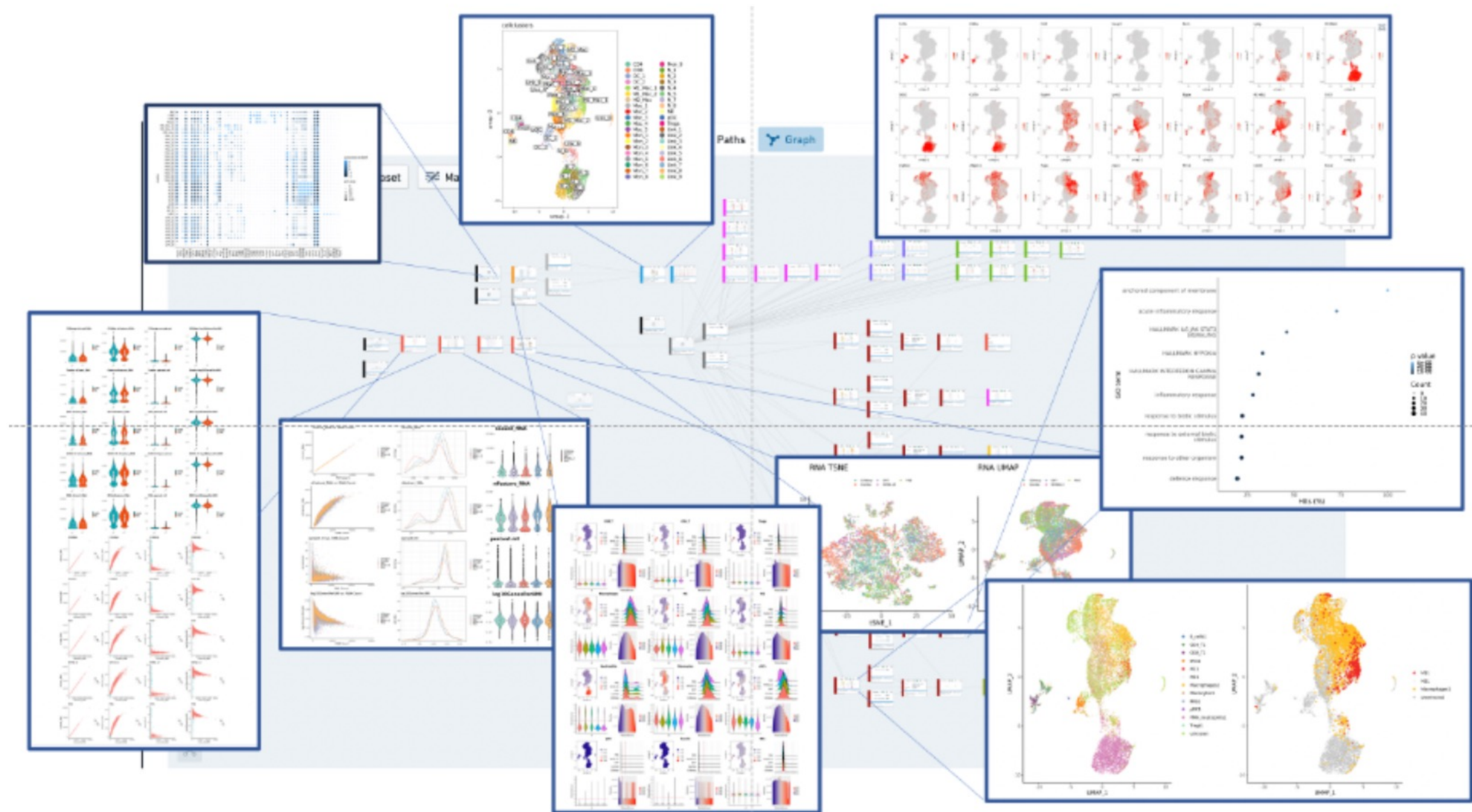
# Cloud Resources

NIDAP – NIH Integrated Data Analysis Platform –bulk and single-cell RNA Seq Analysis Workflows on Palantir Foundry from the CCR Collaborative Bioinformatics Resource (<https://ccbr.ccr.cancer.gov/education-training/nidap-workflows/>)


NCI Cancer Research Data Commons (CBIT) including Cancer Genomics Cloud powered by Seven Bridges (<https://datacommons.cancer.gov/analytical-resource/seven-bridges-cancer-genomics-cloud>)

DNAnexus pilot AWS cloud access with both user-friendly GUI and command line interfaces (send email to [ncibtep@nih.gov](mailto:ncibtep@nih.gov))

# NIDAP

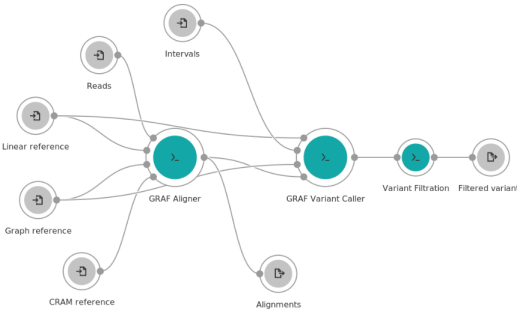


# Cancer Genomics Cloud powered by Seven Bridges

 Public Apps

We offer publicly available Common Workflow Language workflows and tools to enable reproducible bioinformatics.

[Browse 745 apps](#)



**GRAF Germline Variant Detection Workflow**

The GRAF Germline Variant Detection Workflow enables accurate alignment and variant calling by utilizing a genome graph reference that can address the bias and other limitations inherent in linear genome references. Seven Bridges has constructed a comprehensive pan-genome graph that incorporates the...

[Alignment](#) [Variant Calling](#) [Graph](#) [Genomics](#) [CWL1.0](#)

[Run](#)

**RNA-seq alignment - STAR 2.5.4b**

Toolkit version: STAR 2.5.4b

This workflow performs the first step of RNA-seq analysis - alignment to a reference genome and transcriptome. STAR (...)

[Alignment](#) [RNA](#)

[Run](#)

**Whole Exome Sequencing - BWA + GATK 4.0 (with...**

Toolkit version: GATK 4.1.0.0

This Whole Exome Sequencing (WES) workflow identifies variants from a human exome experiment by using the Broad Institut...

[WES\(WXS\)](#)

[Run](#)

**Whole Genome Sequencing - BWA +...**

Toolkit version: GATK 4.1.0.0

This Whole Genome Sequencing (WGS) workflow identifies variants from a human whole-genome resequencing experiment by u...

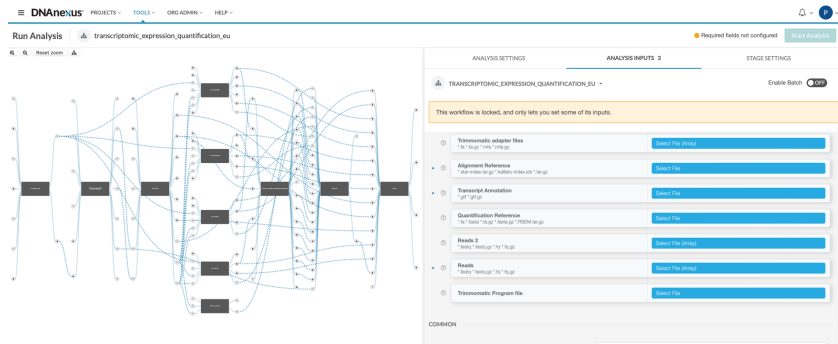
[WGS](#)

[Run](#)



The power of the Cloud via a simple Web-based Interface

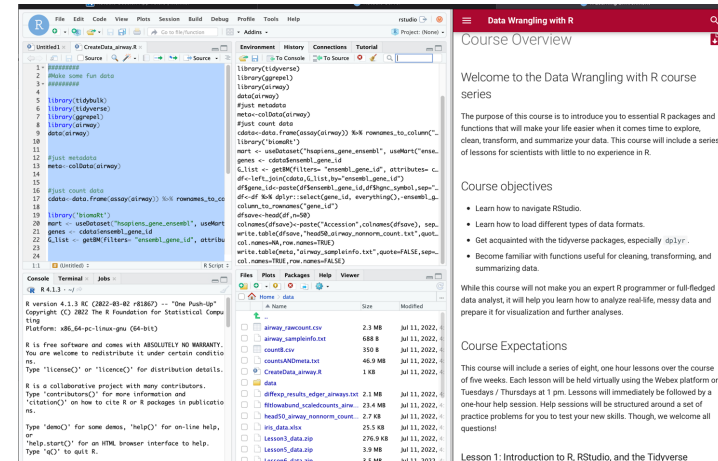
## Pre-built workflows - RNA-Seq Expression Workflow



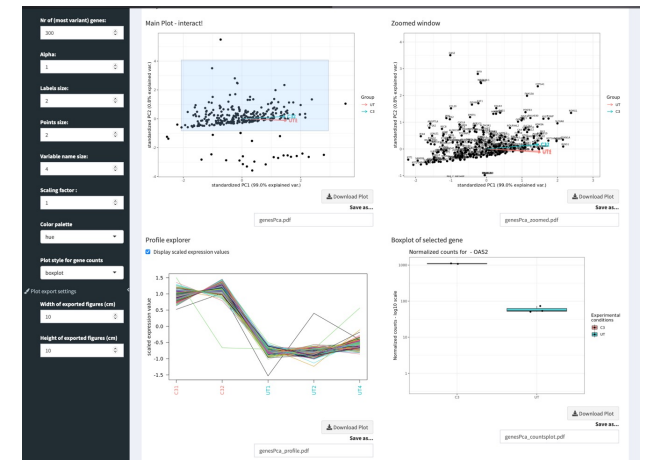
## Custom Solutions - Interactive Data Analysis



## Training Environments - Unix and R



## Hosted R Shiny Apps – PCA Explorer



# Next topic: Core Resources

OSTR | Office of Science and Technology Resources

## Core Resources



Animal Resources



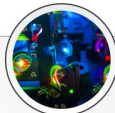
Bioinformatics Biostatistics and  
Computing



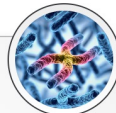
Chemistry and Structural  
Biology



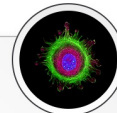
Clinical Research Support



Flow Cytometry



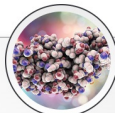
Genetics and Genomics



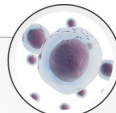
Imaging and Microscopy



Pharmacology



Proteins and Proteomics



Single Cell Analysis

# Focus on: Sequencing Cores

## CCR Sequencing Facility (CCR-SF) (ATRF in Frederick)

- For large scale projects\*
- Short read and long read technologies
- Single cell -omics
- Optical mapping with Bionano

## NCI CCR Single Cell Analysis Facility (SCAF) (Building 41)

- Advanced single cell genomics for researchers on the NIH Bethesda main campus

## NCI CCR Genomics Core (Building 37)

- For small scale projects\*
- Example technologies: short read NGS, long read Oxford Nanopore MinION, Sanger sequencing, digital droplet PCR, etc.

# Other Bioinformatics Resources and Training

All training events are available on the BTEP  
NIH Bioinformatics Calendar

List serv at [list.nih.gov](https://list.nih.gov)

[NIH Library](#) offers training classes in  
software and NGS analyses

Center for Biomedical Informatics and  
Information Technology ([CBIIT](#))

CBIIT NCI Data Science Learning Exchange  
(machine learning, python)

# Bioinformatics List Serves at NIH

Welcome to NIH LISTSERV

[list.nih.gov](http://list.nih.gov)

BIOINFORMATICS-SIG-L (NIH Bioinformatics Special Interest Group)

NIH-DATASCIENCE-L (Data Science at NIH)

SINGLE-CELL-GENOMICS-L (Single-Cell Genomics)

Thank you for your support!

Office of Science and Technology Resources (OSTR)

Dave Goldstein  
Mariam Malik

---

# CCR and NIAID Collaborative Bioinformatics Resources

---

Thank you for helping us by providing expert knowledge about bioinformatics tools and resources

---

We couldn't do the training we do without your support

---

Maggie Cam (CCBR Lead)

---

Parthav Jailwala (CCBR Bioinformatics Manager)

---

Justin Lack (NCBR Lead)

---

A big thank you to all the bioinformatics analysts that have answered questions for us and participated in our "Topics in Bioinformatics" Series

# Genome Analysis Unit (GAU) and BTEP Teams

---

Peter Fitzgerald

---

Carl McIntosh

---

Des Tillo

---

Amy Stonelake

---

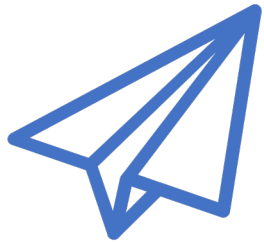
Joe Wu

---

Alex Emmons

Slides and recording from this presentation  
(<https://btep.ccr.cancer.gov/on-line-classes-2022>)

# We want to hear from you



[Email: ncibtep@nih.gov](mailto:ncibtep@nih.gov)



What kind of training is helpful to you?